AI)	

Award Number: DAMD17-02-2-0059

TITLE: Development of a Viral Biological-Threat Bioinformatics

Resource

PRINCIPAL INVESTIGATOR: Elliot J. Lefkowitz, Ph.D.

CONTRACTING ORGANIZATION: The University of Alabama at Birmingham

Birmingham, Alabama 35294-0111

REPORT DATE: October 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command

Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;

Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20040105 185

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188). Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED

October 2003 Annual (30 Sep 2002 - 29 Sep 2003)

4. TITLE AND SUBTITLE

Development of a Viral Biological-Threat Bioinformatics

5. FUNDING NUMBERS
DAMD17-02-2-0059

6. AUTHOR(S)

Elliot J. Lefkowitz, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Alabama at Birmingham Birmingham, Alabama 35294-0111

8. PERFORMING ORGANIZATION REPORT NUMBER

E-Mail: elliotl@uab.edu

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

In response to the potential use of viruses as biological weapons, we have established the Viral Biological-threat Bioinformatics Resource (VBBR) that collects, catalogs, annotates, and analyzes genetic information related to potential viral threats. This work expands upon available knowledge of virus replication, pathogenicity, and virus-host interactions on the basis of individual protein domains, individual genes, and whole genomes. To date, we have constructed a genome and gene sequence database that has been populated with the sequence information for viruses currently listed on the NIH and CDC priority pathogen list. We have also developed a variety of analytical and visualization tools that aid in the analysis of the genomic information coded for by these viruses. Finally, the information developed as a result of this work has been made available to the scientific community through a (currently access-controlled) web site (http://vbbr.genome.uab.edu) that supports research efforts to develop environmental detectors, diagnostics, antiviral compounds, new vaccines, and animal models in support of biodefense research goals.

14. SUBJECT TERMS Biodefense, biological database	terrorism, bioinforma	tics, viruses, sequence	15. NUMBER OF PAGES 13 16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. Z39-18 298-102

Table of Contents

over1	
F 2982	
able of Contents3	
itroduction4	
ody4	
ey Research Accomplishments1	3
eportable Outcomes1	3
onclusions1	3
eferences	
ppendices	

INTRODUCTION:

In response to the potential use of viruses as biological weapons, we have established the Viral Biological-threat Bioinformatics Resource (VBBR) that collects, catalogs, annotates, and analyzes genetic information related to potential viral threats. This work expands upon available knowledge of virus replication, pathogenicity, and virus-host interactions on the basis of individual protein domains, individual genes, and whole genomes. This information has been made available to the scientific community to support current research efforts to develop environmental detectors, diagnostics, antiviral compounds, new vaccines, and animal models.

Specifically we are:

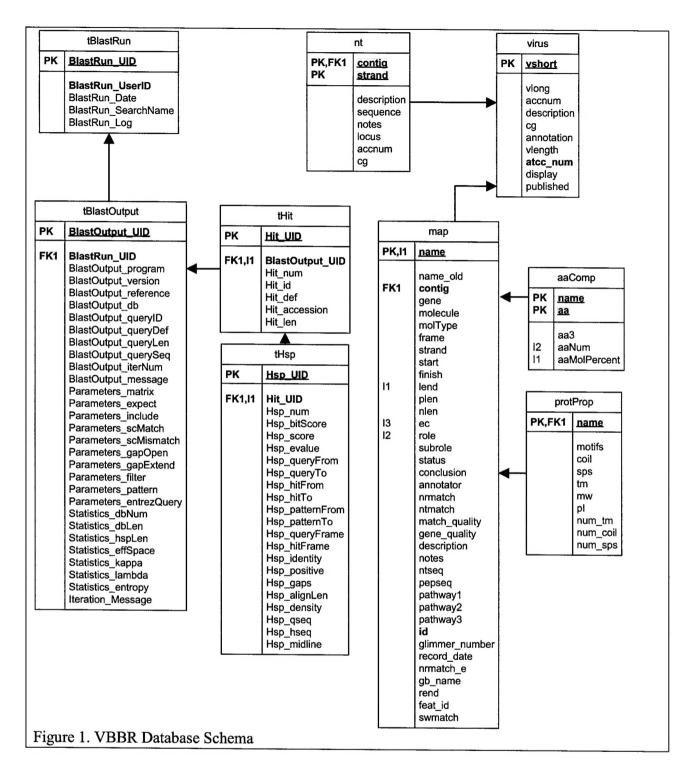
- 1. developing a relational database that supports the data storage, annotation, analysis, and information exchange goals of this proposal;
- 2. developing a world-wide web site that will allow all information and analytical results compiled and generated by this work to be publicized to the scientific community, providing rapid access to specific information as required by an individual researcher;
- 3. collecting existing gene and genomic sequences for viral threat agents and importing them into the database for subsequent annotation and analysis;
- 4. providing computer-automated and human-directed annotation to confirm and update existing biological information in the sequence records for all sequences; and
- 5. performing a variety of analyses on genome and gene sequences to provide additional information on their structure, function, and evolution.

BODY:

Substantial progress has been made in accomplishing many of the tasks outlined in the original statement of work. Accomplishments extend to database construction, data population, development and implementation of analytical and visualization tools, and publication of all available information on a web site. The accomplishments are listed below, itemized according to the original statement of work and task list.

Task 1. To develop a relational database that will support the data storage, annotation, analysis, and information exchange goals of this proposal. (Months 1-6)

• The database has been created based on our previous work on poxviruses and has been updated and refined to better support a wider range of virus threats. An overview of the database schema is provided in figure 1.



Task 2. To develop a world-wide web site that will allow all information and analytical results compiled and generated by this work to be publicized to the scientific community, providing rapid access to specific information as required by an individual researcher. This web site will also provide the necessary security and access controls to ensure that only individuals and groups as designated by the granting agency (USAMRMC) can utilize these resources. (Months 1-12)

- The web site has been established and is accessible at the url: http://vbbr.genome.uab.edu.
- Login to the web site requires registration. Following registration, a user account is established, and access is secured via a username/password. To register, a user needs to

logon to the following web page: https://btd.genome.uab.edu/register/register.asp using the following credentials: Username: genuab; Password: \$mgbf2003 and fill out the form. The user will be vetted for access and when approved will then be notified via Email when the account is established.

• The home page for the VBBR web site is shown below in figure 2.

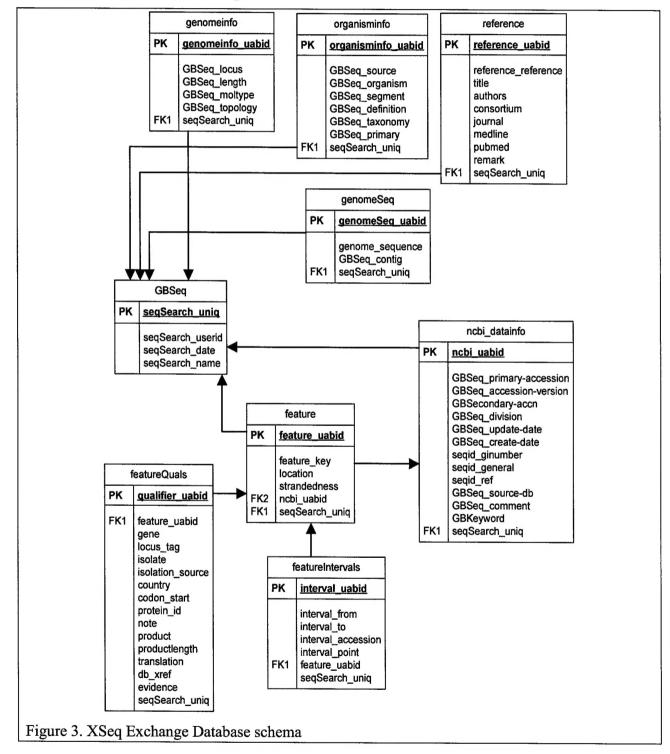
- viril and in the control of the
File Edit View Favorites Tools Help
→ Back ▼ ☆ ヾ ノ ☆ ☆ Search → Favorites → Media ダーム▼ ● 図 ▼ ョ リ → Address ☑ https://btd.genome.uab.edu/vbbr/ マ ぐ Go Links * ラ ▼ CのSnaglt 憧
Address ≥ https://btd.genome.uab.edu/vbbr/
Viral Biodefense Bioinformatic Resource
data analysis reagents literature discussion announcements links
Welcome
Welcome to the Viral Biothreat Bioinformatics Resource Center. This Center has been established to provide resources to the scientific community for basic research and to facilitate the development of novel antiviral therapies and vaccines against virus infections as well as approaches for environmental detection of virions and the rapid diagnosis of disease.
The Center provides a relational database that supports data storage of virus genomic sequences, and annotation and analysis of virus genes; web-based data mining and sequence analysis tools; software for analysis of complete genomes; and a discussion forum.
The Viral Biothreat Bioinformatics Resource Center was created by a consortium of laboratories. An Advisory Committee has been established to help guide the development of this Center. Funding for this project is being provided by the US Army Medical Research and Materiel Command. Suggestions for improvement of the Center
are welcomed (see Feedback). $VBBR$
Last Modified: 08/06/2003 14:06:28
teams & people address book acknowledgements feedback
Copyright © 2002 VBBR Consortium. All rights reserved. Privacy Statement
Si Internet
Figure 2. VBBR Web Site

Task 3. To collect existing gene and genomic sequences for viral threat agents and import them into the database for subsequent annotation and analysis (Months 1-24).

a. We will obtain existing gene and genomic sequences for all viruses that potentially could be used as agents of biological warfare or terror. The

information collected will include the corresponding descriptive annotation provided with the sequences.

An automated parser has been developed (XSeq Exchange) that can parse XMLformatted GenBank sequence records and their annotation and load that information
into the VBBR database. This application will greatly enhance our ability to add
new viral genomes to the VBBR database. The database schema for the parser is
shown below in figure 3.



 The genome and gene database is currently populated with most of the viruses currently on the NIAID A-C priority pathogen list. These are shown on the genomes page of the VBBR web site as displayed in figure 4.

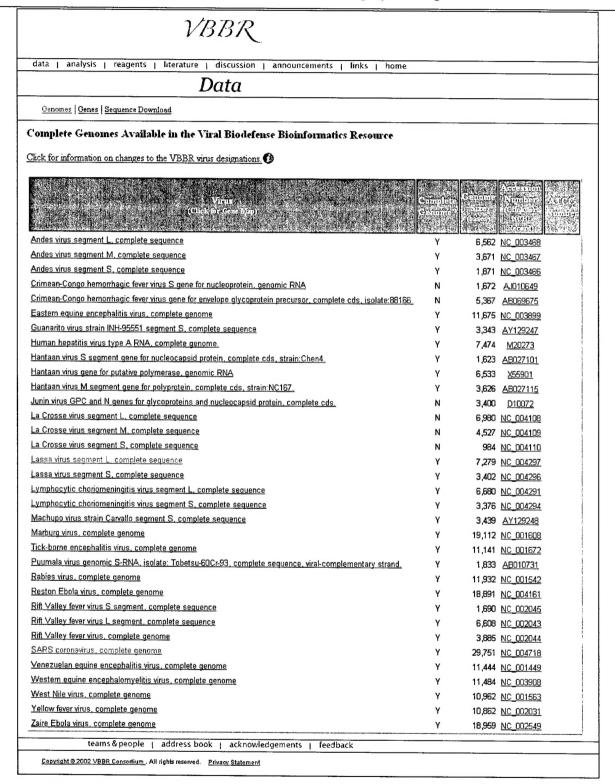


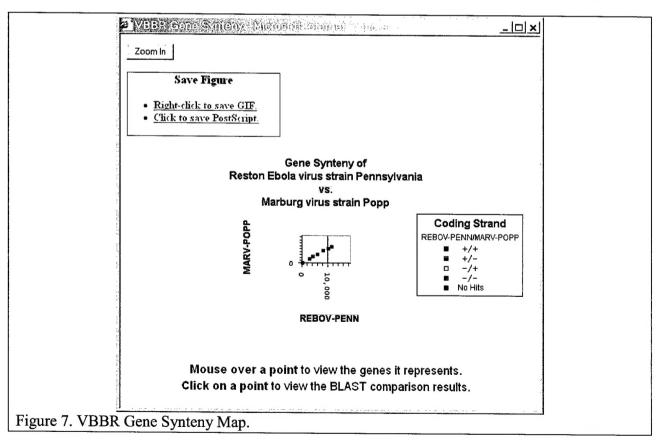
Figure 4. VBBR Genomes Page.

- b. We will also collect and provide the references for publications in the scientific literature that describe studies on the structure and function of orthopoxvirus genes and genomes. We will link this information on our web site to the appropriate gene records in our database.
 - Not yet done. Tasked for year 2.
- c. We will initially obtain viral sequence from public databases such as GenBank. We will also obtain unpublished sequence information through inquires and interactions with other scientists that have been initiated by our collaborators.
 - Not yet done. Tasked for year 2.
- Task 4. To provide computer-automated and human-directed annotation to confirm and update existing biological information in the sequence records for existing as well as newly obtained sequences. (Months 6-24)
 - Computer-automated annotation has been completed for the viruses listed in figure
 Human directed annotation is tasked for year 2. An example annotation record is shown in figure 5.
- Task 5. To perform a variety of analyses on genome and gene sequences to provide additional information on their structure, function, and evolution. (Months 6-24)
 - Initial computer-aided analysis has been completed. Additional analyses tasked for year 2.
- Task 6. To provide through the web site, a set of analytical and visualization tools that will allow users of this resource to mine the data for useful information, to perform comparative analyses between these and other viruses, and to better visualize and assess the significance of their results. (Months 6-24)
 - Currently available analytical and visualization tools include:
 - a. Genome map visualization (Figure 6.)
 - b. Genome protein gene ortholog comparisons
 - c. Gene synteny visualization (Figure 7.)
 - d. BLAST similarity searches
 - e. BLAST search parsing and database storage
 - f. Java-based Visualization of BLAST search results (Figure 8.)
 - A new database and web server has been installed and is now supporting all VBBR resources. This new server provides added security and has greatly increased performance of the VBBR web site.
 - Development of additional analytical and visualization tools are tasked for year 2. These will include tools for creation and visualization of multiple sequence comparisons and tools for the inference and visualization of evolutionary trees.

data i analista i	Diameters				
data analysis reagents		announcements links	home		
	Data				
Genomes Genes Sequence Downle					
BBR Gene Record for: RE	BOV-PENN:RNA-deper	dent RNA polymer			
ast updated by dbo On 7/30/200	3 3:23:53 PM				
ebrane	Marie Constanting of the Constan	, des	Ç.		
10,303		Tran.			18,891
	ola virus strain Pennsylvania				
Genbank Locus Name: RNA-dep					
Gene Name: RNA-dep VBBR Gene Number: REBOV-1	. ,				
Molecule: Protein					
Strand: Forward					
Protein Length: 2212					
Nucleotide Length:					
Start: 11550					
Finish: 18188	2				
Molecular Weight: 252546.6 pl: 8.15	5				
•	endent RNA polymerase [Res	ton Ebola virus]			
in the land	ő Arielysir a later 188				
Amino Acids: Com	osition				
Number of putative Transmembrane Domains: 0 (C	lick for the Complete TransMe	m Record)			
Prosite Motifs: No H	ita.				
Coiled-coil Segments: None					
Secretory Signal Peptides: None		-			
	A Decision and a second				
Database	BLAST Results				
Genbank Protein	View Summary				
Virus Proteins (ortholog search)					
Notes					
socca .					
			1		
			1		
			. i		
Protein Sequence					
	RACG LYSSYSLNPQ LRQCKL	PKHI YRLKFDTIVS KFLS	DTPVAT LPIDYLVPIL	LRSLTGHGDR PLTPTCNO	L DEIINYTLHD
SILGHSQILS VSTAEILING KDIIT	CRENT SLIASIAKLE DVDVSD	PDP SDILKIYNAG DYVI	VHDERI DILGYGDYIF SILGSE GYKIIKYLEP	WKIPLSLLPV TIDGVPHAL LCLAKIQLCS KFTERKGRI	T TONHLSVIND
SFIKENHERS LPHIKDLIWE FYHIN	IPPLF STKVISDLSI FIKDRA	PAVE OTCWDAVFEP NVIG	YNPPNK FSTKRVPEOF	LEGEDFSIES VLNYAGELI	YI SDRNLTPGIN IY LIPONRNFSF
VENUMHYLIP OCYMHVSDYY NPPHN	INISH REYPPEGPSS YRGHIG	SIEG LOOKLWISIS CAOI	SLVEIK TGFKLRSAVM	GDNQCITVLS VFPLKTDPI	E QEQSAEDNAA
KGIDLGQLSL SKPLDYGTIT LTLAV	OVIG GLSFINPEKC FYRNFG	OPVT SGLFQLRVYL ENVN	MKDLFC PLISKNPGNC	SAIDFVINPS GINVPGSQI	OL TSFLRQIVER
MATORTOYPD ARISSPIULD CCDLY AAFLDYYLKA TGAODHITHI ATREK ETLGESQILS YSTAEILINC KDLITT LRELISHREL KDYQOEKIRD HKILL SFIKRNEPS LPHIKDLIVE FYHLN SLIKEKELNIG RITGKLPYLT ROYNWHYLLI OCYMHYSDYY NPERN RVAASLAKVI SACGIFLKED ETFUH KROLDIGOLIS SKPLDYGTI ITLAV SITLIARNKL INTLEHASAD LEDEM OKISCTVOLA QIIREYTUSH IEGER SSALKATEL ASRLTWYTOG SSALKATEL ASRLTWYTOG SSALKATEL ASRLTWYTOG SSALKATEL ASRLTWYTOG SYNSE MAYLYDIRER NINTEDLERIN RAHLH ISDNENSSTD PISSGETREF THEFL LEDAARLEIR TALASFLOPEL TODHNDFSRG DETLIEFYEK FSSNHE	SLIGA TIPCHVEOFK VKWIGO	IEEC BECTURKESN WARS	VAVKDQ VVSAVPNTSR	ISWTIGSGVP YIGSRTEDI	LD HCDQLLADAL (I GQPAIKPRCP
AVALYDIRFR NTNTSDIRHN RAHLH ISDNSNSSTD PISSGETPSE TTUET	TECC TKEVPAQYLT YTSALN	LDLS RYRDNELIYD SNPI	KGGLNC NLTIDSPLVK	GPRINNIEDD LIRFPHIS	W ELAKTYVQSI
LSDAARLFIR TAIASFLOFL KSWII	SUBJECT STREET STREET	INEE THRITCHT AATH	IPKEVS IQNDGHIDLA	ENNYVYNSKS TASNFFHAS	SI AYWRSRKSRK
	(QYLH QLRSMLDTTI YCRFTG	IVSS MHYKLDEVIL EYNS	KKGKFH KWNPCKMIME FDSAIT LAEGEGSGAL	LILOKYSTRI LFINTLATI	H SIESEVVSGF
IQDHNDFSRG DGTLTEPVRK FSSNH LPFFKLGNHN HAHDQDAQEL MNQNI STPRMIIPIM OKUUECOUTU IIANG		JOVE LIMMUALITE NENR	DOTARW AANTITDHID	PUYLKVVVLK VFLSDIEG	LL WINDYLAPIF
IQUENDIFSKE DETLTEPVEK FSSNH LPFFKLGNHN HAHDQDAQEL MUQNI STPRMLLPIM QKVHEGQVTV ILNNS GAGYLIKPIT SSARSSEWYL CLSNL OAFTRDIVID WILMDESDTO TUUR	STNR RSAHQTHKAC LGVIRD	ALQA QVQRGVYVLS HIAQ	YATKNI HCEYIGLGFP	SLEKVLYHRY NLVDTGLG	PL SSVIRHLTNL
LSDARRIFLE TALASFLEYER KSVII. TODHNDFSRG DCTLTEPVRK FSSNH LPFFKLGNHN HAHDODAGET MNON! STPRMLLPIM OKVHEGQVTV ILNNS GAGYLIKPIT SSARSSEWYL CLSNL QAEIRDLYLD YNLWRESRTO TYHFI! FYPEGLIYSN HT	ISTNR RSAHQTHKAC IGVIRD. (TAKG RITKLVNDFL KFSLIV	ALOA OVORGVYVIS HIAO DALK NNSSWYTELK KLPE	YATKNI HCEYIGIGFP VINVCN RFYHTHNCEC	QEKFFVQTLY LQRLRDAE	PL SSVIRHLTNI IK LIERLTGIMR

Figure 5. VBBR Gene Annotation Record.

MEDICAL STREET STREET AND ASSESSED AS A SECOND OF THE SECO	(=1-1
File Edit View Fevorites Tools Help	_ _ _
CHI BOL - B C - C C Absorbe Assert Asserts Asserts Asserts	
Address : , http://btd.genome.usb.edu/vbb/geno_map_al_date.esp 2 ⁻ 66 : Links */	ल ≠ tiCoSnegit ∰
1'BBR	2
data analysis reagents Interature discussion announcements In ks home	
Data	
Can are a Sense Sensence Countries	
Reston Ebola virus strain Pennsylvania	
1 to 18,891 Zoom Level 3 8 Genes	
Remove Labels	
LEGEND Factorina Protein (+ chand) Protein (- chand)	
1	18,891
Note: Overlapping does not represent different reading frames.	
Bank to 3 at Page	
Equily to, Agent Majo Chart	
with the South American	
teams&people address.book acknowledgements feedback	
Gaster Microsoft State Control of Land State St	
Sec	△ 😂 Internet
Figure 6. VBBR Genome Map.	
*	



it About	man of the same of		of distribution the behaviour that goes from the state of
	OHERY	PROGRAM: blastx DATABASE: nr SEQUENCE: sars	
ort By	Quent	BEQUENCE. SAIS	
Bit Score	- 100 to 0	tr = 10 to = 100	MAR 40 : 1777
Query From	✓ 🔛 № 6-100 to 0 № 6-10 to 6-100 .01 to 6-10 Select AI		
Sort :			
it Selection Horizontal Zoom			
Hide Checked Hits Show All Hits	L	evel 1 ·· Level 2 Leve	el 3 🧳 Level 4
4.	Number of Nucleotides		The state of the s
4000 B000	12000 16000 20000	24000 28000	1 29715
gi 26007546 xef	A - area of the Section of Alphabeter and Analysis de Arterior (a. c.)		ORFlab polyprotein [Murine hepatitis vizus]
gi 26008095 ref			transmembrane domain 2 [Bovine coronavirus]
gi 22219303 pdb			Chain A, Structure Of Coronavirus Main Proteinase
gi[25121564]ref			coronavirus nsp3 (MD2) [Murine hepatitis virus]
gi 26008087 ref			coronavirus nsp5 [Bovine coronavirus]
gi 25121551[ref			coronavirus nap6 [Avian infectious bronchitis viru
gi 25121568 ref			coronavirus nsp7 (GFL); growth-factor-like protein
gi 25121569 xef	The second beauty		coronavirus nap9 (RdRp); RNA-dependent RNA polyme
gi[6625761]gb]A	Manager and the second second second		RNA-directed RNA polymerase [murine hepatitis viru
gi 26008891 ref	1972 0.72		coronavirus nspl0 (MB, MTPase/HEL); metal-binding
gi 25121571[ref	EED		coronavirus nspl1 [Murine hepatitis virus]
gi 26008094{ref			coronavirus nsp13 [Bovine coronavirus]
gi 5565844 gb A			spike glycoprotein [murine hepatitis virus]
gi 6706916 gb A			spike glycoprotein [bowine coxonavixus]
gi 6179905 gb A			membrane protein [murine hepatitis virus]
gi 11640712 gb			nucleocapsid protein [Equine coronavirus]
AST Viewer Usage: ICK on HIT NAME to view sequence record. ICK on HIT GRAPHIC to view alignment(s).			
and the second s			

Figure 8. VBBR BLAST Viewer. The figure shows the results of a BLASTX search of the SARS genome against the Genbank non-redundant protein database.

KEY RESEARCH ACCOMPLISHMENTS:

- Establishment of the VBBR Database
- Establishment of the VBBR Web Site
- Development of the XSeq Exchange GenBank sequence parser
- Population of the VBBR database with virus genomic information
- Computational annotation of virus genes
- Implementation of analytical and visualization tools for BLAST searches
- Implementation of analytical and visualization tools for genome ortholog comparisons

REPORTABLE OUTCOMES:

Informatics:

- VBBR Database
- VBBR Web site (http://vbbr.genome.uab.edu)
- VBBR analytical and visualization tools

Presentations:

- USAMRIID/WRAIR/LANL/UAB Workshop, August 11, 2003 WRAIR
- USAMRMC Bioinformatics Workshop, November 4, 2003 (Poster)

Funding proposals submitted extending this work:

NIH/NIAID Proposal in response to the RFP: NIH-NIAID-DMID-04-34,
 "Bioinformatics Resource Centers for Biodefense and Emerging/Re-emerging Infectious Diseases"

CONCLUSIONS:

Development of the Viral Biological-threat Bioinformatics Resource (VBBR) has substantially followed the original task list and essentially all tasks proposed for year 1 have been accomplished. The result is a Bioinformatics Resource Center that can now begin to support the needs of basic and applied biodefense research directed at gaining a better understanding of the pathogenesis, evolution, and overall biology of viral pathogens as well as support the development of detectors, diagnostics, antivirals, and vaccines. Future work will follow the original task list and is aimed at providing human-curated gene records along with additional analytical and visualization tools that can better support understanding the role of individual genes in virus pathogenicity.

REFERENCES:

None other than the VBBR web site.

APPENDICES:

None